

## **Argentina como objeto de investigación. Reflexiones conceptuales y aproximaciones metodológicas para el abordaje bibliométrico de la producción científica sobre temas de alcance nacional**

Sandra Miguel<sup>1,2,3</sup>; Claudia González<sup>1,2</sup>; Mónica Hidalgo<sup>1</sup>

<sup>1</sup>Universidad Nacional de La Plata. Facultad de Humanidades y Cs. de la Educación. Dpto de Bibliotecología, La Plata, Argentina. <sup>2</sup>Instituto de Investigaciones en Humanidades y Ciencias Sociales - IdIHCS (UNLP-CONICET), La Plata, Argentina. <sup>3</sup>SCImago Research Group. [sandra@fcnym.unlp.edu.ar](mailto:sandra@fcnym.unlp.edu.ar)

**Resumen.** En Argentina, las políticas nacionales de investigación de los últimos años manifiestan una clara orientación hacia temas y problemas sociales y productivos locales, en medio de tensiones hacia el interior de la comunidad científica y con los organismos de promoción y evaluación de la investigación entre la pertinencia y la excelencia de los resultados científicos. El proyecto PICT 2011-2183 "Argentina como objeto de estudio en la producción científica con visibilidad regional e internacional. Una aproximación al estudio del dominio temático argentino", acreditado por la ANPCYT, tiene como principales objetivos caracterizar la producción científica con visibilidad en fuentes regionales e internacionales abocada a temas y problemáticas circunscritas al dominio temático "Argentina"; y analizar y describir el perfil de los investigadores y la composición de las redes de colaboración que conforman las comunidades científicas que estudian estos temas. Esta ponencia tiene por objeto presentar avances de este proyecto de investigación y reflexionar acerca de los problemas conceptuales y metodológicos que plantea el abordaje de un estudio bibliométrico de esta naturaleza basado en la información presente en bases de datos bibliográficas y en las opciones de recuperación y descarga de la información que ofrecen.

### **1. Introducción**

Diversos estudios encontrados en la literatura se han dedicado al análisis bibliométrico de la producción científica del dominio geográfico argentino. Algunos a nivel multidisciplinar (Quesada Allue y Gitlin, 1995), (Fernández, 1998), (Fernández, 1996), (Miguel, 2008, 2009), otros sobre áreas temáticas específicas (Rabinovich, 1992) y (Ribichich, 1998) en ecología; (Merlino Santesteban, 2007) en matemática; y en áreas prioritarias (Merlino Santesteban, 2013); (Oesterheld, 2002) y (Blake et al., 2002) en agronomía; (Liberatore, 2008) en psicología. También han sido objeto de estudio dominios institucionales, especialmente universidades nacionales, por ser instituciones que desempeñan un protagónico rol en la producción de conocimientos científicos (Herrero-Solana, 2001), (Argañaraz, Sosa y Rapela, 2003), (Miguel, Moya-Anegón y Herrero-Solana, 2006). Con diferentes alcances temporales y/o disciplinares todos estos estudios centran su atención en la producción científica desarrollada en Argentina,

utilizando como criterio de demarcación el país de residencia de los investigadores mencionado en los datos de afiliación institucional de las publicaciones. Esto permite tener aproximaciones sobre el perfil y las tendencias de la investigación que se realiza en el país, y contar con indicadores bibliométricos que aportan información relativa a la producción científica nacional.

El proyecto PICT 2011-2183 *Argentina como objeto de estudio en la producción científica con visibilidad regional e internacional. Una aproximación al estudio del dominio temático argentino* se ha planteado como uno de sus objetivos el caracterizar la producción científica abocada a temas y problemáticas circunscritas al dominio temático “Argentina”. Esto significa que, a diferencia de los estudios que analizan la producción científica “del país”, el presente proyecto centra la mirada en la investigación que se realiza “sobre el país”, es decir donde Argentina sea objeto de estudio o lo sea la delimitación geográfica que refiere al alcance territorial de las investigaciones. Mientras los primeros estudian la producción desde el lugar de afiliación de alguno de los autores de las publicaciones científicas, en el segundo caso, se plantea trabajar con la literatura que hace referencia explícita al nombre del país o de algún lugar geográfico comprendido en el territorio nacional en los títulos, palabras claves y/o resumen de los artículos publicados en revistas visibles en fuentes bibliográficas regionales e internacionales.

La temática y objetivo del proyecto tienen especial relevancia en el contexto de las actuales políticas nacionales de promoción científica, en las cuales se manifiesta un claro interés en la producción de conocimientos orientados a la resolución de problemas sociales, productivos y medioambientales de carácter local. Diversos indicios de ello se pueden encontrar en los planes estratégicos de ciencia, tecnología e innovación del Ministerio de Ciencia, Tecnología e Innovación Productiva de la Nación de los últimos años, como el “Plan Bicentenario” 2006-2010 (MINCYT, 2006), y el Plan Argentina Innovadora 2020 (MINCyT, 2011). En el primero se hace especial énfasis en la necesidad de complementar la estrategia tradicional de promoción de la investigación por disciplinas, con el fomento de la producción de conocimientos cuyo principal eje sean los problemas y oportunidades existentes en la sociedad y en el medio productivo. En el segundo, la orientación es hacia el desarrollo de innovaciones que redunden en

una mejora de la calidad de vida de la población y en un incremento del valor agregado de los bienes y servicios producidos en el país.

La agenda científica además define temáticas consideradas áreas problema-oportunidad y áreas temáticas prioritarias. De este modo se promocionan y financian proyectos de investigación con determinados perfiles y orientaciones en temas de relevancia nacional. La orientación de recursos hacia áreas prioritarias cuenta con una larga trayectoria en diversos países, aunque no por ello resulta de fácil definición e implementación. En general, las áreas suelen definirse a partir de diagnósticos previos y de una selección de campos relevantes que ya cuentan con cierta trayectoria; pero además, por el análisis de su potencial aplicabilidad. También podrían definirse a partir de la identificación de áreas de vacancia (Unzué, 2012). Cómo obtener luego indicadores de resultados de investigación en esas áreas es otro asunto complejo, debido a que la información requerida para estos fines no está disponible directamente en las fuentes de datos actualmente disponibles, y los sistemas de organización y clasificación de la información no son compatibles con las demandas de datos para el cálculo de tales indicadores (Vila Seoane, Arber y Bassotti, 2011).

Otra cuestión es el análisis de los modos de producción de conocimiento que se caracteriza por una tendencia de transición de las investigaciones realizadas individualmente a las basadas en los grupos de investigación y en las redes de colaboración (Katz y Martin, 1997). Este fenómeno que ha dado lugar a una importante revalorización del rol de los recursos humanos y de los vínculos que se establecen entre ellos, está siendo considerado como la base del capital científico social de las instituciones y de los países, favoreciendo el desarrollo de redes de conocimiento (Bozeman, 2005). Estas redes se conciben como configuraciones en las que interactúan actores de diversas procedencias, con el fin de abordar problemas concretos y proponer soluciones complementando sus conocimientos y capacidades. En este escenario, el segundo objetivo general que se plantea este proyecto es analizar y describir las redes de colaboración y la composición de los grupos de investigación que abordan temas y problemáticas locales, así como interpretar las estrategias seguidas por los investigadores en la conformación de esos grupos y redes.

Todo ello en pos de contribuir a una mayor comprensión de la dinámica de la producción de conocimientos que gira en torno a temas de alcance local, y favorecer la

retroalimentación de información en base a evidencias empíricas sobre el estado de desarrollo de estos temas en la literatura científica, así como para la promoción de estrategias de integración e intercambio entre los grupos y redes que participan en su abordaje.

## **2. Alcances y limitaciones del abordaje del tema-problema mediante el método bibliométrico**

Existen diferentes abordajes teóricos y metodológicos que intentan describir, explicar e interpretar el complejo y multidimensional fenómeno de la dinámica de la producción del conocimiento científico, no existiendo un único enfoque que lo pueda abordar completamente. Entre los diferentes abordajes metodológicos, la aproximación bibliométrica y sociométrica de la literatura científica centra su atención en el análisis del tamaño, crecimiento y distribución de la bibliografía científica en diferentes tipos y niveles de agregación temáticos, institucionales, geográficos, etc. (Okubo, 1997) por una parte, y en el estudio de la estructura social de los grupos que la producen y la utilizan por otra (López Piñero, 1972). Si bien esta aproximación existe desde hace tiempo, los avances en las actuales capacidades computacionales han facilitado enormemente el análisis de conjuntos de datos a gran escala, y los progresos recientes en las técnicas de visualización hacen posible mostrar múltiples aspectos de la actividad de manera estructurada y sintética; y permiten además la representación de fenómenos complejos que facilitan la realización de posteriores análisis en profundidad desde diferentes perspectivas.

Pese a este potencial de la bibliometría, su elección como método supone la asunción de ciertas limitaciones en el alcance de las dimensiones y complejidad del fenómeno estudiado, ya que la investigación no siempre proyecta resultados tangibles y las publicaciones científicas son sólo una dimensión más dentro del quehacer y del devenir del sistema científico (Moravcsik, 1989).

Considerando estas ventajas y limitaciones el proyecto plantea realizar una investigación exploratoria-descriptiva respondiendo a un diseño semiestructurado y a un planteo inductivo basado en el análisis bibliométrico y sociométrico de las publicaciones científicas, complementado con la consulta a documentos oficiales y

entrevistas a investigadores.

La elección de las fuentes de datos a utilizar para realizar los estudios bibliométricos es otro asunto controvertido. En este proyecto se decidió basar los análisis a partir de la literatura científica registrada en tres bases de datos. Por un lado, Scopus, de Elsevier, por su amplia cobertura de revistas internacionales y su alcance multidisciplinar. Por otro, SciELO y RedaLyC por ser las fuentes de difusión de la literatura científica más importantes a nivel latinoamericano, y que recogen la producción publicada en revistas nacionales y regionales que siguen estándares de calidad editorial.

El período temporal de cobertura del estudio es desde 2008 a 2012.

## **2.1 instrumentación de las búsquedas**

### **2.1.1 Recuperación del conjunto de registros**

#### **Estrategia de la búsqueda**

La primera cuestión conceptual a resolver en torno al planteamiento de las consultas en las fuentes de datos consistió en definir la estrategia de la búsqueda. Este paso supuso reflexiones y definiciones acerca de cómo traducir, de manera exhaustiva pero a la vez precisa, la necesidad de información (los trabajos donde Argentina sea el objeto de estudio o el alcance territorial de las temáticas investigadas) al lenguaje y opciones de búsqueda ofrecidas por las interfaces de cada una de las tres bases de datos seleccionadas.

Sabido es que no siempre se menciona en el título, palabras clave o resumen la palabra “Argentina” o el nombre de alguna de sus provincias. Por tanto, no resulta sencillo saber en qué casos un trabajo refiere a un tema de alcance nacional / local sin estos delimitadores. Pero, la única manera de identificar los trabajos desde la perspectiva de la recuperación de información en bases de datos bibliográficas es a través de términos explícitos presentes en los campos de las bases de datos. Se decidió entonces que Argentina era uno de los términos clave para la búsqueda y que su presencia en los campos de título, palabras clave o resumen era indicio de su uso como tema o alcance territorial de la investigación abordada. Sin embargo, ese solo término no se consideró suficiente para representar el conjunto esperado. Una exploración preliminar de las

bases de datos permitió identificar que muchos trabajos hacen mención explícita a nombres de provincias argentinas pero no mencionan el país en ninguno de los tres campos. Por tanto se decidió plantear una estrategia de búsqueda que incluyera, como términos, tanto el nombre del país como de los de cada una de las provincias argentinas. Esto no significó haber resuelto y contemplado todas las posibilidades terminológicas que pudieran referir a estudios con alcance nacional; pero al menos permitía recuperar un conjunto de documentos relevantes basado en criterios objetivos y repetibles en cualquier otro estudio posterior.

La ecuación de búsqueda planteada no supuso condiciones con relación al país de afiliación de los autores de los trabajos. Esto significa la posibilidad de recuperar estudios realizados en Argentina o fuera de ella. Justamente, y aunque es de esperar que haya una mayor proporción de publicaciones de alcance nacional que sean resultados de investigaciones desarrolladas en el país, este proyecto también se propone identificar aquella porción de trabajos que sobre Argentina realizan investigadores de instituciones del extranjero.

Desde el punto de vista metodológico y excepto en el caso de Scopus -que permite plantear una estrategia de búsqueda compleja simultáneamente en los campos de título, palabras clave y resumen-, en los casos de SciELO y RedALyC el planteamiento de una consulta de esta naturaleza es una limitación. En el primer caso (SciELO) porque no permite la búsqueda simultánea en los tres campos clave seleccionados, y por tanto hay que realizarla usando la opción de búsqueda “todos los índices”. De otro modo, repetir las búsquedas en cada uno de los campos separadamente, genera duplicaciones y solapamientos en los resultados. RedALyC, además de lo anterior, no permite efectuar búsquedas con operadores booleanos que combinen los términos, además de que la lentitud del sistema torna imposible cualquier intento de recuperación de información con los requerimientos planteados.

### **Exportación de resultados de búsqueda**

A estas dificultades se suman las limitaciones en la exportación de los resultados de las búsquedas.

En el caso de Scopus si bien presenta muy ricas y variadas posibilidades de exportación en diferentes formatos, limita la cantidad de resultados exportables a 2000 registros por

vez. Considerando que la cantidad de registros recuperados conforme a la estrategia y período temporal (aprox. 20.000) es ampliamente superior a esa cifra, se tuvieron que buscar alternativas para realizar las búsquedas en forma parcial y exportar en bloques de registros que no superen la cantidad máxima permitida. La estrategia adoptada fue filtrar los resultados por año combinando el ordenamiento de los mismos desde los más nuevos a los más antiguos y viceversa para realizar las exportaciones múltiples que permitieran descargar, finalmente, la totalidad de los registros recuperados e importarlos a otra base de datos *ad-hoc*.

En el caso de SciELO hubo que realizar las búsquedas en los portales de cada país de manera independiente (y no así en el portal SciELO regional), ya que solo de esa manera es posible descargar los registros en un formato que permita la reutilización de los datos para análisis. Este formato es ISO 2709. Los registros exportados se importaron en una base de datos Winisis, y luego haciendo uso de lenguaje de formateo de campos se volvieron a exportar en un archivo de texto plano separado por comas, legible por un programa de planilla de cálculo.

RedALyC no ofrece opciones de exportación, por lo que resulta imposible descargar registros resultado de búsquedas. Por tanto, en este caso se optó por solicitar a los desarrolladores del portal una bajada en formato CSV de los registros que cumplieran con la ecuación de búsqueda definida para los objetivos del proyecto. Se obtuvo el listado de los registros aunque parcialmente, ya que los datos no incluyen nombres de los autores de los trabajos ni las instituciones de afiliación de los mismos.

### **Relevancia de los resultados de búsqueda**

Las primeras revisiones sobre los resultados de las búsquedas mostraron la presencia de registros que cumplían con los requerimientos planteados en la ecuación, pero que no eran relevantes a los objetivos del proyecto. Este ruido documental fue motivado principalmente por nombres homónimos a los de las provincias argentinas pero que refieren a lugares, instituciones, apellidos de personas que nada tienen que ver con las temáticas de alcance nacional. Reformular la estrategia de búsqueda quitando los nombres de las provincias no suponía una solución posible, no encontrando manera de resolver esta cuestión desde la perspectiva conceptual e instrumental que ofrecen los

sistemas de recuperación de información de las fuentes utilizadas (aunque la limitación no esté en el contenido de fuentes sino en el motor y lenguaje de recuperación de los sistemas).

## **2.2. Tratamiento de los datos**

### **2.2.1. Registros no relevantes**

El descarte de registros no relevantes recuperados supone una tarea de inspección cualitativa que no es posible realizar con métodos automáticos (al menos no con las herramientas tecnológicas disponibles por/para el proyecto). Esto significa afectación de recursos humanos para la revisión registro por registro para determinar su pertinencia o no a los objetivos del estudio. En el caso de Scopus se estima un descarte de registros no relevantes del 30% aprox., en Scielo este porcentaje se eleva al 50%, y en Redalyc aún no se ha podido estimar pero se espera un porcentaje similar.

Un aspecto a considerar, y no de menor importancia, es la dificultad para determinar qué registros son pertinentes temáticamente, ya que incluso la presencia de los términos buscados no siempre es garantía de que los temas tratados sean los que se plantea el estudio. En otras palabras resulta muy difícil planificar una depuración de registros no pertinentes mediante métodos objetivos y automáticos, por la ambigüedad del lenguaje y los diferentes contextos en los que uno u otro término puede ser empleado en el título, palabra clave o resumen de una publicación científica.

### **2.2.2. Estructuración de los datos**

Del modo en que se pudieron exportar / descargar los registros de las diferentes bases de datos no en todos los casos resulta posible disponer de los mismos de manera óptima para su tratamiento y posterior análisis. Las fuentes bibliográficas organizan los datos en campos con fines de recuperación de información y no con fines de recuentos estadísticos; aún las fuentes que como Scopus y *Web of Science* ofrecen información bibliométrica. Esta limitación de las estructuras de las bases de datos bibliográficas para fines métricos ya ha sido señalada hace varios años por Godin (2000), y más recientemente por Ferrara y Salini (2012) quienes plantean que el complejo y creciente escenario de los estudios bibliométricos hace necesario pensar nuevos modelos conceptuales basados en un enfoque multidimensional.



Entre las formas que adoptan los registros bibliográficos y las necesidades de organización y estructuración para fines bibliométricos no existen herramientas de software que resuelvan todos los problemas y necesidades. Sin embargo hay algunos desarrollos que son gratuitos y que pueden utilizarse, como por ejemplo el programa de conversión de registros procedentes de Scopus a una base de datos de estructura cuasi-relacional desarrollado por Leydesdorff en 2008 y que está disponible en <http://www.leydesdorff.net/software/scopus/>. Esta herramienta se ha probado y si bien es robusta y permite la importación de grandes cantidades de datos, presenta algunas limitaciones entre las se pueden mencionar: no importa las palabras clave; algunos títulos de revistas se importan de manera incompleta; no importa los ISSN de las revistas, etc. Asimismo, Scopus presenta problemas de compatibilidad con códigos de caracteres que se producen en casos de presencia de subíndices y superíndices, que la exportación basada en UTF-8 corte los registros y los divide en varios con campos en blanco.

A los registros procedentes de las otras fuentes utilizadas solo se los pudo estructurar en filas y columnas en una planilla de cálculo y no presentan facilidades para realizar recuentos de algunos datos como autores, por ejemplo, que aparecen uno seguido del otro separados con punto y coma al estilo de referencia bibliográfica. Por otro lado, los títulos de las revistas se hallan abreviados, y así otras cuestiones de índole similar que dificultan los procesos de unificación y normalización de datos con fines de recuentos estadísticos.

### **2.2.3. Registros procedentes de diferentes fuentes de datos**

Unificar los registros procedentes de las tres bases de datos sería ideal. Sin embargo requiere también una tarea laboriosa y manual de compatibilización de datos con diferentes estructuras además de procesos de normalización y descarte de duplicados. En este último caso, si bien podrían plantearse algunas fórmulas para facilitar su identificación, no siempre resultan satisfactorias porque los títulos de los artículos, que son los elementos clave para establecer los patrones de diferenciación de registros, en algunas fuentes están en inglés (por ejemplo en Scopus) y en otras en español (por ejemplo en SciELO) aún cuando el texto del documento se encuentre en español. Tampoco los títulos de las revistas suelen estar normalizados y los ISSN pueden variar

según se trate del número asignado a las revistas en soporte impreso, digital o mixto.

#### **2.2.4. Cualificación de registros y desarrollo de una fuente de datos enriquecida**

Los diferentes sistemas de clasificación que emplean las fuentes de datos supone la necesidad de una homologación entre esquemas que no siempre puede realizarse de manera sencilla y confiable. Por lo general, son las revistas y no los artículos los que están clasificados temáticamente. Tal es el caso de Scopus, SciELO y RedALYC y de la mayoría de las fuentes bibliográficas. En el nivel de los artículos las descripciones temáticas son por lo general palabras clave de autor, que si bien tienen la ventaja de aportar una mayor especificad, no suelen estar normalizadas.

Por tanto, clasificar temáticamente en el nivel de artículo como estrategia de enriquecimiento cualitativo de los registros objeto de análisis implica una tarea que requiere experticia y no está exenta de subjetividad. Teniendo en cuenta que el proyecto involucra todas las áreas del conocimiento esta tarea resulta mucho más compleja además de costosa. A ello se suma la clasificación geográfica a nivel de provincias que permita luego realizar estudios por regiones. Si bien esto estará dado a partir de la presencia de los términos en el título, palabras clave o resumen, no es posible determinar a priori en qué campo serán encontrados esos cualificadores, debiendo reunir esa información de manera bien estructurada para permitir realizar los recuentos estadísticos.

Y por último, la falta de normalización de los datos, tanto procedentes de una sola fuente como de la integración de varias fuentes, es una problemática siempre vigente. En este proyecto se requiere normalizar instituciones, países y autores, además de la clasificación temática y geográfica que es otra forma de normalizar datos asignados.

#### **2.3. Asociación de la producción a las áreas prioritarias definidas por MINCyT**

La identificación de registros que abordan áreas prioritarias es otra problemática, ya que como se señalara en estudios previos (Vila Seoane, Arber y Bassotti, 2011) no es directa la traducción entre el lenguaje que define esas áreas y las palabras clave presentes en las bases de datos bibliográficas u otras fuentes.

El MINCyT ha definido seis áreas como prioritarias, Agroindustria, Ambiente y Desarrollo sustentable, Desarrollo social, Energía, Industria y Salud. Dentro de ellas,

produce un desglose de hasta 34 subáreas entre las que se encuentran denominaciones tan específicas como “Autopartes” y otras tan generales como “Habitat”. Cada subárea posee una breve nota de alcance. Así, por ejemplo, para la subárea “Mejoramiento de cultivos y producción de semillas” la nota de alcance indica “Incorporación de valor a los granos a partir del mejoramiento genético para generar una fuerte industria nacional de semillas que abastezca a la producción de los cultivos mayores y secundarios y genere exportaciones de mayor contenido tecnológico.”

El enriquecimiento de la fuente de datos con esta clasificación es de vital importancia para darle mayor pertinencia a esta investigación. La formación bibliotecaria y la experiencia profesional de la mayoría de los miembros del equipo (o de personal tercerizado de similar perfil) debiera ser suficiente para emprender esta tarea de clasificación. Se propone emplear una estrategia que a partir de la clasificación de la totalidad de los registros en las seis grandes áreas propuestas por la Clasificación de Areas Científicas y Tecnológicas de la OCDE (2006), se realice una equivalencia con las áreas prioritarias principales, para que luego cada clasificador trabaje con subgrupos de registros afines temáticamente.

Tabla 1

<b>Clasif. OCDE (2006)</b>	<b>Áreas prioritarias MINCyT</b>
Cs. Naturales y Exactas	Ambiente y desarrollo sustentable
Ing. y Tecnologías	Industria/Energía
Cs. Médicas y de la Salud	Salud
Cs. Agrícolas	Agroindustria
Cs. Sociales	Desarrollo social
Humanidades	

El segundo paso en la estrategia es verificar qué subáreas son las más prioritarias. Por ejemplo, si se profundiza en un segundo nivel de correspondencia, el gran área del

conocimiento Cs. Naturales y Exactas se divide en 7 subdisciplinas, de las cuales una de ellas es Ciencias de la Tierra y Relacionadas con el Medio Ambiente. Las siete subáreas correspondientes al área prioritaria Ambiente y Desarrollo sustentable, caen allí. Por ello, se supone que una vez que uno cuenta con la clasificación general de los registros por área y subárea de conocimiento, selecciones de subconjuntos de registros *ad-hoc* permitirán a los clasificadores desarrollar áreas de expertitud que redunden en un desarrollo más ágil en la preparación de la fuente de datos, a la vez que posibilitará la obtención de resultados parciales sin tener que esperar a que el proceso se complete totalmente.

### **3. Discusión y conclusiones**

Varias reflexiones e interpretaciones surgen para la discusión. Por un lado decir que, en gran medida las limitaciones encontradas para el abordaje del tema-problema objeto del proyecto de investigación que se analiza en esta ponencia, no radica en limitaciones del método bibliométrico en sí mismo, sino más bien en dificultades conceptuales y metodológicas inherentes al lenguaje, a las fuentes de datos, a la recuperación y tratamiento posterior de la información recuperada.

Por un lado se encuentran las dificultades conceptuales para definir el conjunto de publicaciones cuyos temas sean de “alcance nacional”. La cuestión aquí es determinar qué es el alcance nacional de una investigación? ¿Cómo se define y delimita conceptualmente? ¿Cómo se representa en el lenguaje empleado en una publicación científica? Y ¿Cómo se traduce en el lenguaje de recuperación de información de los sistemas bibliográficos?. Por otro lado, cabe preguntarse cómo se vinculan con esos temas aquellas temáticas de “relevancia nacional” o áreas prioritarias” en la agenda científica del país? Puesto que como ya se ha advertido tampoco resultan fáciles de delimitar. Los diferentes lenguajes: natural-documental, político-científico y las diferentes lenguas, principalmente, español-inglés en las que los investigadores argentinos comunican los resultados de investigación generan dificultades de traducción que dificulta la etapa de la definición de las estrategias de búsqueda.

La aproximación propuesta en este proyecto si bien no salva todas las limitaciones, al menos permite realizar un primer acercamiento al estudio del conjunto de las publicaciones que hacen explícita la delimitación geográfica de las investigaciones ya

sea con fines de hacer un recorte temático o territorial.

Otras de las dificultades encontradas tiene que ver con las diferentes formas en que pueden plantearse las estrategias de búsquedas en los sistemas. No todos operan con las mismas interfaces dificultando el planteo de las consultas, como también de las descargas de registros en un único formato pasible de procesamiento automático.

La eficiencia de los sistemas de información para la obtención de resultados relevantes de una búsqueda ha sido y sigue siendo un problema de la recuperación de información. La presencia de ruido documental motivada por registros relevantes desde el punto de vista objetivo (es decir, que responden a la ecuación de la búsqueda) pero no pertinentes a la necesidad (objetivos del proyecto), es un asunto de difícil resolución, al menos en el caso del problema-tema que aquí se pretende abordar. Considerando que la cantidad de registros recuperados es elevada, la revisión de los mismos para desechar los no pertinentes supone un costo en tiempo y recursos humanos que se traduce en un problema/limitación.

Por otro lado, los diversos esquemas de clasificación utilizados por las distintas fuentes, sumado a la falta de normalización de datos de autores, afiliaciones institucionales, e incluso de los nombres de las revistas, son otros problemas no menores, especialmente para estudios de nivel macro y multidisciplinarios como el que plantea este proyecto, no solo por la cantidad de registros objeto de análisis sino por su diversidad.

A modo de conclusión agregar que no es habitual encontrar en la literatura el planteo de estos inconvenientes del abordaje de estudios bibliométricos como un problema. Sin embargo, el análisis pormenorizado de estas cuestiones debería ser objeto de más reflexiones que permitan desarrollar propuestas de mejora de los sistemas a fin de minimizar las limitaciones señaladas en los estudios métricos. Los problemas aquí planteados son en su mayoría de pertinencia del campo de la bibliotecología y ciencia de la información, ya que están directamente vinculados con la representación, organización, recuperación y tratamiento de datos bibliográficos para fines bibliométricos.

## **Bibliografía**

- Argañaraz, Eugenia Bustos; Sosa, Alicia Centeno y Rapela, María Virginia. 2003. Análisis bibliométrico de la producción científica de los investigadores

con proyectos aprobados por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Córdoba: 1996-1999. Transinformação [en línea], vol. 15, n° 2, p.231-243. [Citado 29 Jul 2013]. Disponible en World Wide Web: <http://www.brapci.ufpr.br/download.php?dd0=16743>

- Blake, R. et al. 2002. Las ciencias agropecuarias en la Argentina. Ciencia Hoy, vol. 12, n° 70, p.31-51.
- Bozeman, B. 2005. La problemática contemporánea de la aplicación e institucionalización de la evaluación de la ciencia, la tecnología y la innovación. SECTIP Taller "Evaluación de Resultados e Impacto de la Ciencia, la Tecnología y la Innovación" . Buenos Aires: Ministerio de Educación, Ciencia y Tecnología. Secretaría de Ciencia, Tecnología e Innovación Productiva. [en línea]. [Citado 23 Jul 2013]. Disponible en World Wide Web: [http://www.secyt.gov.ar/publicaciones/Taller\\_evaluacion\\_impactos\\_cyt.pdf](http://www.secyt.gov.ar/publicaciones/Taller_evaluacion_impactos_cyt.pdf)
- Fernández, Raúl. A. Estudio bibliométrico de la producción científica. La Plata: Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, 1998.
- Fernández, Raúl A. Estudio bibliométrico preliminar de la producción científica Argentina. Análisis cuantitativo de artículos publicados. Albornoz, Mario; Kreimer, Pablo, and Glavich, Eduardo, compiladores. Ciencia y sociedad en América Latina. Bernal, Buenos Aires : Universidad Nacional de Quilmes 1996, p. 138-150.
- Ferrara, A. y Salini, S. 2012. Ten challenges in modeling bibliographic data for bibliometric analysis. Scientometrics, vol. 93, n° 3, p. 765-785 [doi:10.1007/s11192-012-0810-x](https://doi.org/10.1007/s11192-012-0810-x)
- Godin, B. 2000. Outline for a History of Science Measurement. Science, Technology, & Human Values [en línea], vol. 27, n° 1, pp. 3-27. Montreal, Canadá. [Citado 27 Jul 2013]. Disponible en World Wide Web: [http://www.csiic.ca/PDF/Godin\\_1.pdf](http://www.csiic.ca/PDF/Godin_1.pdf)
- Herrero-Solana, Victor. 2001. Producción científica de la Universidad Nacional de Mar del Plata : Análisis de dominio. Nexos, vol. 8, n° 14, p. 4-10.
- Katz, J.S. y Martin, B.R. 1997. What is research collaboration?. Research Policy, vol. 26, p. 1-18.

- Liberatore, Gustavo; Hermosilla, Ana. 2008. La producción científica argentina en psicología: un análisis de la visibilidad e impacto en el ámbito internacional y su comparación con países de la región. Revista Interamericana de Psicología [en línea], vol. 42, n° 3, p. 507-512. [Citado 29 Jul 2013]. Disponible en World Wide Web: <http://www.redalyc.org/articulo.oa?id=28442310>
- López Piñero, J. M. 1972. El análisis estadístico y sociométrico de la literatura científica. Valencia : Facultad de Medicina. Centro de Documentación e Informática Médica.
- Merlino Santesteban, Cristian. 2013. Desempeño científico argentino en cinco áreas prioritarias de I+D. Una mirada a través del SCImago Journal & Country Rank Revista Iberoamericana de Ciencia, Tecnología y Sociedad, vol. 8, n° 22, p. 33-56.
- Merlino Santesteban, Cristian. 2007. Investigación matemática argentina recogida en MathSci (2000-2005). Información, Cultura y Sociedad, vol. 16, p. 86-105.
- Miguel, Sandra. 2008. Aproximación cienciométrica al análisis y visualización del dominio científico argentino, 1990-2005. Granada : Universidad de Granada. Tesis de doctorado.
- Miguel, Sandra; Moya-Anegón, Félix de y Herrero-Solana, Víctor. 2006. Aproximación metodológica para la identificación del perfil y patrones de colaboración científica de dominios científicos universitarios. Revista Española de Documentación Científica, vol. 29, n° 1, p. 36-55.
- Miguel, Sandra y Moya-Anegón, Félix de. 2009. La ciencia argentina bajo la lupa de los indicadores cienciométricos. Una mirada crítica de la actividad científica argentina. La Plata : Ediciones Al Margen.
- MINCYT. 2006. Plan Estratégico Nacional de Ciencia, Tecnología e Innovación “Bicentenario” (2006-2010). Buenos Aires : Ministerio de Ciencia, Tecnología e Innovación Productiva, noviembre de 2006. [en línea]. [Citado 30 Mar 2010]. Disponible en World Wide Web: [http://www.mincyt.gov.ar/plan\\_bicentenario/documentos\\_finales/plan\\_bicentenario\\_publicacion.pdf](http://www.mincyt.gov.ar/plan_bicentenario/documentos_finales/plan_bicentenario_publicacion.pdf)

- MINCYT. 2011. Argentina Innovadora 2020: Plan Nacional de Ciencia, Tecnología e Innovación. Buenos Aires: Ministerio de Ciencia, Tecnología e Innovación Productiva. [en línea]. [Citado 30 Jul 2013]. Disponible en World Wide Web: [http://www.mincyt.gob.ar/publicaciones#anc\\_8063](http://www.mincyt.gob.ar/publicaciones#anc_8063)
- Moravcsik, M. J. 1989. Cómo evaluar la Ciencia y a los científicos. Revista Española de Documentación Científica, Vol. 12, n° 3, p. 313-325.
- Oesterheld, Martín; Semmartin, María; Hall, Antonio. 2002. Análisis bibliográfico de la investigación agronómica en la Argentina. Ciencia Hoy [en línea], vol. 12, n° 70, p. 52-62. [Citado 29 Jul 2013]. Disponible en World Wide Web: <http://www.agro.uba.ar/users/semmarti/papers%20pdfs/agronomia%20ciencia%20hoy.pdf>
- Unzué, M. 2012. Universidad, políticas públicas y áreas prioritarias en Ciencias Sociales. Revista de la Facultad de Ciencias Sociales de la Universidad de Buenos Aires, n° 80.
- Okubo, Y. 1997. Bibliometric Indicators and Analysis of Research Systems: Methods and Examples. (Report No. STI Working Papers 1997/1). Paris : Organisation for Economic Co-operation and Development .
- Quesada Allue, L. A. y Gitlin, D. S. 1995. Scientific output in Argentina 1966-1983. Scientometrics, vol. 34, n° 1, p. 27-35.
- Rabinovich, J. E. 1992. Publications of scientists of developing countries in international journals: are they channels to the international circuit for colleagues that only publish in national journals? Scientometrics, vol. 23, n° 1, p. 91-103.
- Ribichich, Alejandra M. y López de Casenave, Javier. 1998. La ecología en Argentina: ¿Qué y cómo investigan los ecólogos argentinos? Mitos y realidades entre el atraso y el desarrollo. Interciencia, vol. 23, n° 5, p. 280-285.
- Vila Seoane M.F.; Arber, G. y Bassotti, F. 2011 Palabras clave para la búsqueda de información en áreas prioritarias. En: Albornoz M., y Paza L. Agenda 2011: temas de indicadores de ciencia y tecnología. 1ra. ed. Buenos Aires : Red Iberoamericana de Indicadores de Ciencia y Tecnología.